ARI Research Note 92-51

AD-A253 387

Meta Analysis of Aircraft Pilot Selection Measures

David R. Hunter
U.S. Army Research Institute

Eugene F. Burke
United Kingdom Ministry of Defence

ARI AVSCOM Element
Michael Benedict, Chief

MANPRINT Division
Robin L. Keesee, Director

June 1992

92-18823

ARMY RESEARCH INSTITUTE

United States Army
Research Institute for the Behavioral and Social Sciences

92

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

MICHAEL D. SHALER
COL, AR
Commanding

Technical review by

Charles A. Gainer
Gabriel P. Intano
John E. Stewart
Dennis C. Wightman

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1992, June	Final Oct 88 - Jan 91

4. TITLE AND SUBTITLE
Meta Analysis of Aircraft Pilot Selection Measures

5. FUNDING NUMBERS
63007A
793
1210
H1

6. AUTHOR(S)
Hunter, David R. (ARI); Burke, Eugene F. (Ministry of Defense, UK)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
ARI AVSCOM Element
4300 Goodfellow Boulevard
St. Louis, MO 63120-1798

8. PERFORMING ORGANIZATION REPORT NUMBER
14 ARI Research Note 92-51

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: PERI-S
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
---

11. SUPPLEMENTARY NOTES
--

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
distribution is unlimited.

12b. DISTRIBUTION CODE
--

13. ABSTRACT (*Maximum 200 words*)

For this research, the meta-analytic procedures described by Hunter and Schmidt in 1990 were applied to a database of 476 correlations based on an overlapping sample of 432,324 cases. These correlations were obtained from a review of the research literature on aircrew selection published from 1920 to 1990. Over 200 studies that dealt with aircrew selection were identified. Of that number, 69 reported correlations between some independent measure and a pilot training performance criterion. Analyses were conducted of the overall aggregated set of correlations and subsets selected on the basis of date of study, type of predictor measure, type of aircraft, and sample characteristics. These analyses showed a decline in the mean validity correlations obtained over the previous 50 years. In addition, differences in the mean correlations were observed among the various types of predictor measures. In general, job sample measures were the best predictors of performance, followed by psychomotor coordination and biographical inventories. Possible applications of the results in the interpretation of previous research and in the design of future research are discussed.

14. SUBJECT TERMS
Pilots — Personnel selection
Aircrew — Meta analysis
Selection — Test validation

15. NUMBER OF PAGES
40

16. PRICE CODE
--

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)

FOREWORD

This report continues the extensive analysis of the aircrew selection literature previously documented in a narrative review and an annotated bibliography. In this study, meta analysis is used to quantitatively integrate 50 years of research that spans multiple military services and nations.

The results of this analysis point to an overall decline in the validity of pilot selection measures but at the same time firmly establish the validity of a number of measures being used operationally or being investigated. This study also supports claims for validity generalization of selection measures by showing equivalent validities for a number of measures across services, nations, and aircraft.

This research was conducted in the U.S. Army Research Institute for the Behavioral and Social Sciences MANPRINT Division by the Aviation Systems Command Element of the U.S. Army Research Institute Fort Rucker Field Unit in collaboration with Science 3 (Air), United Kingdom Ministry of Defence.

The information contained in this report was briefed to the NATO AGARD Committee on Pilot Selection and also to program management personnel of the Aviation Systems Command. The report will be made available to other researchers in the field of aircrew selection and will be used to focus Army research to improve aviation system designs through better specification of the abilities and attributes of aircrew members.

META ANALYSIS OF AIRCRAFT PILOT SELECTION MEASURES

EXECUTIVE SUMMARY

Requirement:

The purpose of this study is to evaluate various measures for the prediction of performance in pilot training.

Procedure:

A search of the computer databases and a manual search of armed service bibliographies, Psychological Index, and reference lists of all citations was conducted. The criterion for inclusion was the description of some process or measure being used or being considered for use for aircrew selection or classification. This criterion was loosely applied, however, to obtain a thorough representation of the available literature. Aircrew in this case refers primarily to pilots, although some studies dealing with navigators were included.

The database that resulted from this search is described in Hunter and Burke (1990). From that database, all studies reporting predictive validities for aircraft pilots were identified. The correlation values, sample size, and other information regarding the characteristics of the sample and the study were coded and recorded for analysis.

The meta-analytic procedures described by Hunter and Schmidt (1990) were applied to the database to generate mean correlations and variances for the overall set of correlations and specified subgroups.

Findings:

Over 200 studies dealing with aircrew selection were located. Of those studies, 69 contained correlations between some independent measure and a pilot performance criterion. A total of 476 individual correlations, based on an overlapping sample of 432,324 cases, were used in the analyses.

Analyses were conducted of the overall set of correlations and subsets selected on the basis of date of study, type of predictor measure, type of aircraft, and sample characteristics.

These analyses revealed a declining mean correlation over the previous 50 years. In addition, differences in the mean correlations were observed among the various types of predictor measures. In general, job sample measures were the best predictors of performance, followed by psychomotor coordination and biographical inventories. Age is negatively related to performance (older trainees have the least likelihood of completing training), while personality measures consistently are the least related to performance.

Utilization of Findings:

The results of this research can be used to better interpret the findings of previous research in aircrew selection and guide in the choice of measures used for operational pilot selection. In addition, these results should shape future research efforts through delineation of the relationships between predictor measures and training criteria more stable than those obtained in single studies.

META ANALYSIS OF AIRCRAFT PILOT SELECTION MEASURES

CONTENTS

LIST OF TABLES

LIST OF FIGURES

META ANALYSIS OF AIRCRAFT PILOT SELECTION MEASURES

Introduction

The training of aviators is an expensive and lengthy process for the military services. Training courses are typically about one year in length with from 150 to 250 hours of flight time, at a cost ranging from $500 to over $3,000 per flight hour. The high cost of training makes failure to complete training especially alarming. For the United States Air Force, estimates of the typical cost of a failure during pilot training range from $50,000 (Hunter, 1989) to $80,000 (Siem, Carretta, & Mercatante, 1987). These figures are probably typical of those for aviators from most air forces and navies, with the cost of failures for army aviators being somewhat less due to the lower cost of the predominately helicopter-based training.

The high cost of training and training failures, coupled with a training attrition rate that has historically been in the range of 20 to 40 percent (with the notable exception of recent US Army attrition rates of approximately 10 percent), have provided the stimulus for a great deal of military research on the aviator selection process. The history of this research is described by Hunter (1989) in a narrative review of the literature.

While the narrative review technique can provide a general description of what research has transpired, it does not provide a methodology for the efficient integration of disparate research findings. Fortunately, such a methodology is now at hand in the form of meta analysis (Hunter, Schmidt, & Jackson, 1982; Hunter & Schmidt, 1990). Meta analysis provides the means for cumulating and integrating research findings from multiple studies. In the case at hand, this technique will allow for the development of a single best estimate of the correlation between some predictor measure and a criterion (flying training). From these estimates of the population correlations, comparisons may be made of the validities of specific predictor measures or classes of measures. Thus, one may ask whether, based upon fifty years of cumulated research studies, one measure is superior to another for the prediction of flying training performance. Or, one may observe whether one class of predictors (such as psychomotor coordination tests) is superior to another class of predictors (such as personality measures).

From these comparisons, conclusions may be drawn regarding the likely optimal composition of batteries for aircrew selection, and the most promising areas for research in predictor measure development may be identified. In addition, because of the nature of the database that will be used in this study, inferences may be made regarding the validity generalization of measures across applicant populations, military services, and types of aircraft.

The object of this study, then, is to develop a database of studies that will support the aims listed above and to apply the techniques of meta analysis to that database so as to be able to make comparisons between and among individual predictor measures and classes of measures. The specific areas of interest are (a) validities of classes of predictor/selection measures, (b) validities for specific aircraft, nationalities and services, (c) generalizability of validities across groups and aircraft, and (d) validities of specific predictor measures.

Sample

The sample for this study consisted of all studies on aircrew selection published circa 1920 to 1990. A thorough review of the literature cited in Psychological Abstracts along with United States and British military reports was conducted to identify relevant studies. The results of that search of the literature are provided as an annotated bibliography in Hunter & Burke (1990).

From the collection of all studies dealing with aircrew selection, those studies that reported correlations between one or more predictor measure and an aircrew performance measure were identified. There were 69 such studies, with a total of 664 correlations. These correlations constituted the sample used in this study. The citations for the studies containing these correlations are given in Appendix A.

Procedure

Each study was reviewed, and the correlation, sample size, and certain other information (given in Table 1) regarding the study were coded and recorded in a database. The predictor measures were classified following the system described by Hunter (1989) for the General Category (Table 2) and the system described by Pearlman (1979) for the further breakdown of the general cognitive measures into more specific categories (Table 3).

In some cases, several correlations were reported in a single study for a particular measurement instrument. For example, a study might report several correlations between measures taken in a flight simulator for a group of individuals and subsequent performance in flight training. In those cases where multiple measures and correlations were reported for a logically single instrument, the correlations were averaged using the Fisher Z transformation to produce a single validity correlation. This process reduced the sample of correlations from 664 to 476. The distribution of study characteristics for these correlations is given in Table 4. While the sample is predominately based upon general cognitive measures taken from United States Air Force personnel undergoing fixed-wing training, correlated with a dichotomous (pass/fail) criterion, other sources of data also make a substantial contribution. While there is not enough data to allow a complete factorial evaluation of every combination of study characteristic, in many cases there are enough data points (correlations) to allow for meaningful analyses.

These correlations are conceptually, but not necessarily statistically independent (Schmitt, Gooding, Noe, & Kirsch, 1984), as some studies reported correlations for logically independent measures (for example, psychomotor coordination and arithmetic reasoning) based upon measurements taken from a single group of individuals.

Finally, the signs of error-scored measures (e.g., psychomotor coordination) were reflected so that a positive correlation indicated that superior test performance is associated with superior performance in training. An exception to this treatment, however, was that given to personality measures. Because there was no a priori expectation regarding the direction of prediction of these measures (for example, should one expect superior flying performance to be associated with high or low authoritarianism) their signs were not changed. This conservative treatment assumes an underlying population validity of zero, with the observed dispersion of positive and negative correlations being a random process. Additional research might address alternative treatments of this problem.

Table 1

Study information recorded in the database.

Author(s) Name(s)
Date
Name of Predictor Measure
General Category of Predictor Measure
Specific Category of General Cognitive Predictor Measure
Sample Size (N)
Correlation
P-Q Split (proportion in pass/fail categories for dichotomous criterion)
Criterion Category
Sample Description
Sample Nationality
Sample Service
Aircraft Type

Table 2

Predictor Measure General Categories.

General Cognitive
Personality
Information Processing
Job Sample
Biographical Inventories
Psychomotor Coordination
Composites/Batteries
Other

Table 3

Predictor Measure Specific Categories.

General Intellect
Verbal Ability
Quantitative Ability
Spatial Ability
Perceptual Speed
Manual Dexterity
Reaction Time
Mechanical Ability
Aviation Information
General Information
Education *
Age *

* Included in Other category from Table 1; all others included in the General Cognitive category.

Table 4

Distribution of Study Characteristics

Predictor Measure Category	Number of Correlations	Sample Size
General Cognitive	218	250,212
Personality	50	23,889
Information Processing	28	13,072
Job Sample	16	2,822
Biographical Inventories	22	27,962
Psychomotor Coordination	73	42,893
Composites/Batteries	34	35,589
Other	35	35,885
Sample Service	**Number of Correlations**	**Sample Size**
Air Force	286	335,850
Navy	127	72,905
Army	36	19,944
Civilian	27	3,625
Sample Nationality	**Number of Correlations**	**Sample Size**
United States	366	403,453
United Kingdom	24	3,445
Canada	52	9,743
Other	34	15,683
Aircraft Type	**Number of Correlations**	**Sample Size**
Fixed Wing	416	408,516
Rotary Wing	60	23,808
Criterion Category	**Number of Correlations**	**Sample Size**
Dichotomous (Pass/Fail)	404	400,201
Continuous	72	32,123
Total	476	432,324

Data Analysis

Hunter & Schmidt (1990; Table 3.1) list 11 possible study artifacts which will alter the values of outcome values and for which corrections are sometimes possible in meta analysis. These range from sampling error (which may be addressed with meta analysis) to variance due to extraneous factors (which is not addressed by meta analysis). This study attempts to correct only for the most basic of these artifacts--sampling error--and will therefore, constitute what Hunter & Schmidt call a "bare bones" meta analysis.

Sampling error, the variability of study results associated with departures from population correlation values due to random effects associated with the choice and size of the sample upon which the correlation values are based, is also cited by Hunter & Schmidt as the principal cause of variability in study results. Therefore, while this is a "bare bones" meta analysis, the results should still account for a majority of the explainable variance in the research findings.

The basic process for the meta analysis is the computation of a mean correlation from the individual study correlations. The correction for sampling error amounts to weighting each study correlation by its associated sample size. The formula used (from Hunter & Schmidt; 1990, page 100) is:

$$\bar{r} = \frac{\Sigma\,[\,N_i\, r_i\,]}{\Sigma\, N_i}$$

where r_i is the correlation in study i and N_i is the number of persons in study i. The variance of the correlations is similarly weighted and is computed as:

$$\delta_r^2 = \frac{\Sigma\,[\,N_i\,(\,r_i - \bar{r}\,)^2\,]}{\Sigma\, N_i}$$

The variance attributed to sampling error is computed as:

$$\delta_e^2 = \frac{(\,1 - \bar{r}^2\,)^2}{(\,\bar{N} - 1\,)}$$

From these two values, one may obtain the estimate of the variance of the population correlations as:

$$\delta_p^2 = \delta_r^2 - \delta_e^2$$

(Hunter & Schmidt; 1990, page 109)

These equations were implemented in the dBase III command language (See Appendix B) and used to compute mean correlations and associated variances for various groupings of correlations. In addition, the proportion of variance remaining unexplained after reduction for sampling error was computed and reported.

The analyses were conducted in a hierarchical sequence--beginning with all correlations combined and subsequently disaggregating the correlations based upon the study characteristics of interest (e.g., predictor measure, sample nationality, etc.). For each analysis, the mean correlation and three variances (observed, error, and true or corrected) were computed, along with the percentage of unexplained variance (the ratio of true to observed). For those cases in which a negative true variance was calculated, the variance was taken to be zero.

Results

Historical trends

The cautions voiced by Hunter & Schmidt (1990) regarding the over interpretation of results from aggregated higher level analyses should be heeded in the review of these results. Analyses of heterogeneous samples of correlations in which the study characteristics are markedly different can be accused of making apples-and-oranges comparisons. This caution notwithstanding, let us point out that there are many instances in which apples and oranges are indeed combined; for example, under the heading of fruit. Just as a decline in fruit production over the last 50 years would be of interest, so also should be a decline in the validity of predictor measures.

Table 5

Historical Distribution of Validities

Decade	Number of r	Total Sample	Mean Sample	Mean r	δ_r^2	δ_e^2	δ_p^2
1941 - 1950	80	158,516	1,981	.2470	.0128	.0004	.0124
1951 - 1960	103	130,273	1,265	.2377	.0172	.0007	.0165
1961 - 1970	104	41,828	402	.1428	.0137	.0024	.0113
1971 - 1980	78	15,534	199	.1197	.0074	.0049	.0025
1981 - 1990	111	86,173	776	.0852	.0152	.0013	.0139

As Table 5 (and Figure 1) shows, there is a definite downward trend in the mean validities obtained over the last 50 years. Even disregarding the decade from 1941-1950 during which many of the large-scale studies from World War II were conducted, the decline is still evident. Several explanations for this decline suggest themselves: (a) Attenuation of the variability of the applicant pool; (b) Movement toward more extreme P/Q splits in the dichotomous criterion (proportions in the fail and pass groups moving away from an optimal 50/50 distribution); and, (c) Changes in the nature of training. However, the present data do not provide an adequate basis for explanation for this observation, which must be left, for now, to conjecture.

To investigate whether this decline would hold up with disaggregated data, two additional analyses were conducted: one using all correlations for USAF fixed-wing training and a pass/fail criterion, and one using only the general ability predictor measures for the same group. The results from these analyses are shown in Figures 2 and 3 respectively. Both these analyses show the same pattern of decline in validity as the combined data. Further disaggregation to specific predictor measures was not possible because of limited data.

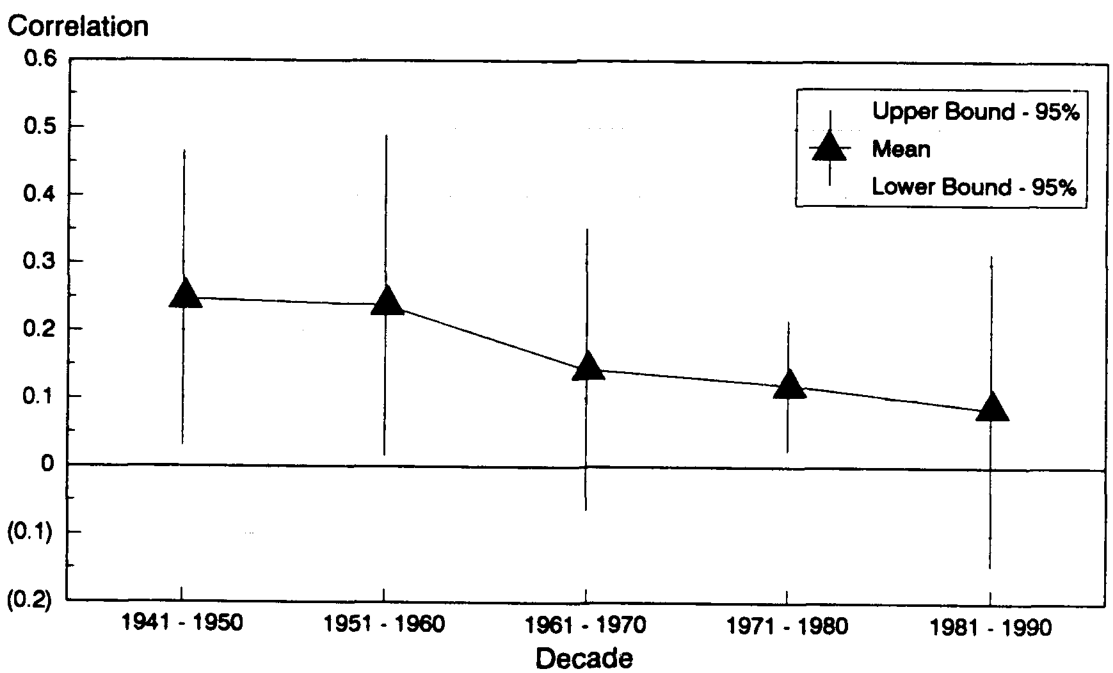

Figure 1. Historical trend in validity.

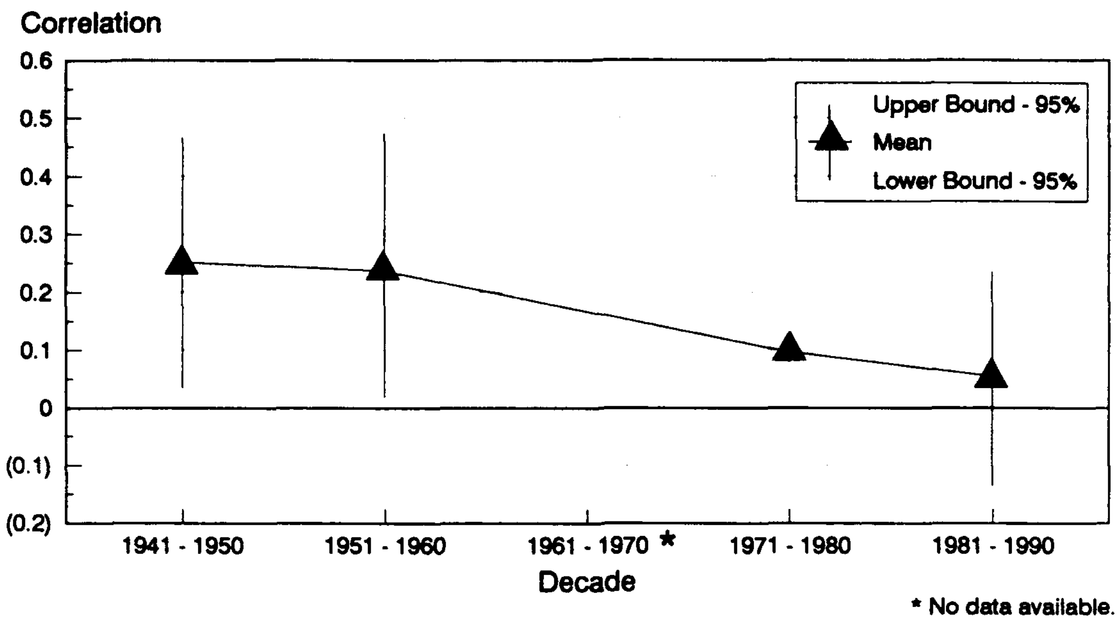

Figure 2. Historical trend in US Air Force validity: Fixed-wing, pass/fail.

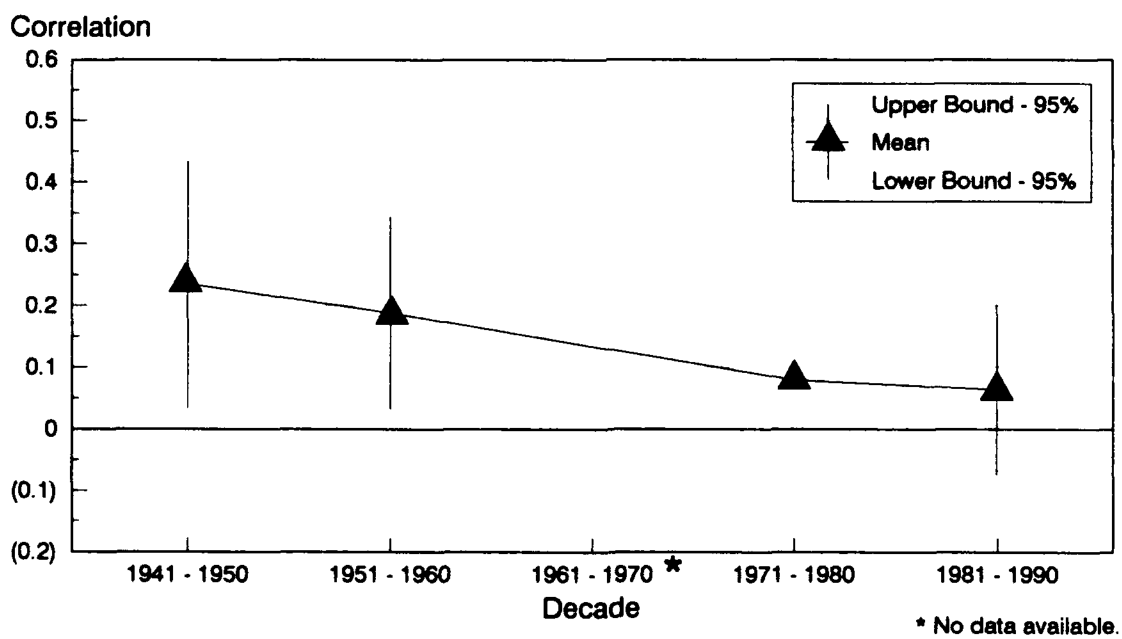

Figure 3. Historical trend in US Air Force validity: Fixed-wing, pass/fail, general ability.

Predictor Measures

The analyses of the predictor measures (at the general category level) are presented in Table 6. The best predictor of pilot performance was the job sample measure, followed by measures of psychomotor coordination and biographical inventories. The least predictive were the personality measures, with a mean correlation of .1168. The standard deviation of the personality measure validities is .1349, giving us a 95% confidence interval for this validity of +/- .2644. This interval includes zero; therefore, we would be unable to reject the null hypothesis of zero validity for this set of measures. (Subject to the note given earlier about the treatment of signs for this category of measures.) Similar evaluations of the other predictor measure sets could be performed, using the corrected estimate of variance, from which the variance due to sampling error has been removed.

The categories of Composite/Battery and Other are included in this analysis solely for the sake of completeness of reporting. The validities reported for the Composite/Battery category are for scores derived from the combination of a number of separate tests. For example, this category includes correlations between the US Navy's flight aptitude rating and training performance, where the flight aptitude rating is a combination of several measures, including written tests and subjective evaluations. The Other category includes validities for measures such as age, physical fitness, and education. Some of these measures are broken out and analyzed separately in the analysis of specific predictor measures. However, in the present instance, these categories represent the leaves and tree bark of our apples-and-oranges analysis and as such should not be interpreted as having any special meaning.

Table 6. Validity Coefficients as a Function of Predictor Type

Predictor Measure	Number of r	Total Sample	Mean r	δ_r^2	δ_e^2	δ_p^2	Percent Unexp.
General Cog	218	250,212	.1924	.0119	.0008	.0111	93
Personality	50	23,889	.1168	.0202	.0020	.0182	89
Info Process	28	13,072	.2256	.0176	.0019	.0159	89
Job Sample	16	2,822	.3272	.0150	.0045	.0105	70
Bio Inventory	22	27,962	.2646	.0109	.0007	.0102	94
Psych Coord	73	42,893	.3035	.0129	.0014	.0115	89
Comp/Battery	34	35,589	.1934	.0228	.0009	.0219	96
Other	35	35,885	.0889	.0424	.0010	.0414	98
Total	476	432,324	.1973	.0189	.0010	.0179	95

Sample Service

Table 7 shows the overall validities for each of the military services (for all nations) and for those studies which used civilian student pilots. The proportion of variance in the validities which is associated with sampling error for these groups is relatively small compared to the proportion remaining (81 to 96%). Considering the heterogeneity of these groups, such a level of unexplained variability is not unexpected, and indicates the need for further disaggregation. Although the mean validities vary among the groups, the differences are not great considering the variances.

Table 7

Validity Coefficients as a Function of Sample Service

Sample Service	Number of r	Total Sample	Mean r	δ_r^2	δ_e^2	δ_p^2	Percent Unexp.
Air Force	286	335,850	.2061	.0185	.0008	.0177	96
Navy	127	72,905	.1697	.0178	.0016	.0161	91
Army	36	19,944	.1546	.0208	.0017	.0190	92
Civilian	27	3,625	.1701	.0368	.0071	.0298	81

Sample Nationality

There were no substantial differences in validities among the nations represented in this sample. However, the variance did differ, with the unexplained variance for the United Kingdom being smaller than that of the United States, Canada, or the Other nations.

Table 8

Validity Coefficients as a Function of Sample Nationality

Sample Nationality	Number of r	Total Sample	Mean r	δ_r^2	δ_e^2	δ_p^2	Percent Unexp.
United States	366	403,453	.1997	.0187	.0008	.0179	96
United Kingdom	24	3,445	.1880	.0181	.0065	.0116	64
Canada	52	9,743	.1715	.0312	.0051	.0261	84
Other	34	15,683	.1541	.0132	.0021	.0112	84

Aircraft Type

The validities for Fixed-Wing aircraft, as shown in Table 9, are slightly, but not significantly, higher than those for Rotary-Wing aircraft. In both cases, however, the amount of unexplained variance is still substantial.

Table 9

Validity Coefficients as a Function of Aircraft Type

Aircraft Type	Number of r	Total Sample	Mean r	δ_r^2	δ_e^2	δ_p^2	Percent Unexp.
Fixed Wing	416	408,516	.1998	.0188	.0009	.0179	95
Rotary Wing	60	23,808	.1545	.0184	.0024	.0159	87

Criterion

Because artificially making a dichotomy out of an otherwise continuous variable (such as flying performance) acts to attenuate the validities with predictor measures, one might have expected to observe a higher mean correlation for the validities which used a continuous criterion as compared to those which used a dichotomous criterion. Such is not the case for these data, however. Although the 95% confidence intervals for these two validities overlap, and hence are not significantly different, nevertheless the direction of difference is contrary to expectation. Whether this is a chance fluctuation or it is telling us something about (possibly) the quality or reliability of the continuous performance indexes used in these studies cannot be readily determined from these data. However, the data are intriguing and possibly deserve additional research.

Table 10

Validity Coefficients as a Function of Criterion Type

Criterion Type	Number of r	Total Sample	Mean r	δ_r^2	δ_e^2	δ_p^2	Percent Unexp.
Dichotomous	404	400,201	.2021	.0184	.0009	.0175	95
Continuous	72	32,123	.1378	.0211	.0022	.0190	90

Predictor-Study Characteristic Relationships

Analyses up to this point have been at the uppermost level of aggregation. With the data in Table 11 begins the process of disaggregation into meaningful subgroups with, hopefully, a reduction in the proportion in unexplained variance.

Since the primary study characteristic of interest is the predictor measure, these analyses are built around that element. Table 11 reports the mean validities for each of the general predictor measure categories for each of the military services and civilian samples. It is at this point that empty cells begin to appear, in which fewer than three validities were found.

For the Air Force (all nations) subsample, the relative ordering of predictor measures is much the same as for the combined, aggregate sample. Job sample measures are the best predictors, followed by psychomotor coordination and biographical inventories. In addition, there is a reduction in the percentage of unexplained variance for several of the predictor measures. In particular, the variance of the biographical inventory validities is now quite low (.0005), although 61% of the variance is still unaccounted for.

For the Navy subsample, there were not enough validities for either the job sample or psychomotor coordination measures to compute mean validities. The best single measure for this group is the biographical inventory, although the variance of these validities (.0203) is far greater than that of the Air Force group.

Lacking sufficient data on job sample or biographical inventory measures, the best single predictor for the Army subsample is psychomotor coordination. For the military subgroups, then, there is a consistent ordering, when the data are available, of the best predictor measures. This is not the case for the civilian subgroup, however, where the best predictor is the information processing measure.

Table 11

Average Validity Coefficients for Predictor-Sample Service Combinations

Predictor Measure	Number of r	Total Sample	Mean r	δ_r^2	δ_e^2	δ_p^2	Percent Unexp.
			Air Force				
General Cog	152	218,459	.1919	.0120	.0006	.0113	95
Personality	27	15,619	.1442	.0223	.0017	.0206	93
Info Process	13	9,569	.2566	.0133	.0012	.0121	91
Job Sample	13	2,172	.3243	.0194	.0048	.0147	75
Bio Inventory	6	15,129	.2875	.0009	.0003	.0005	61
Psych Coord	60	38,525	.3090	.0136	.0013	.0123	91
Comp/Battery	7	17,457	.2150	.0321	.0004	.0317	99
Other	8	18,920	.0994	.0583	.0004	.0578	99
Total	286	335,850	.2061	.0185	.0008	.0177	96
			Navy				
General Cog	49	29,298	.1955	.0109	.0015	.0093	86
Personality	14	6,890	.0712	.0100	.0020	.0080	80
Info Process	10	2,926	.1038	.0065	.0034	.0032	49
Job Sample	1	196	--	--	--	--	--
Bio Inventory	15	12,796	.2475	.0213	.0010	.0203	95
Psych Coord	2	344	--	--	--	--	--
Comp/Battery	15	7,656	.1435	.0235	.0019	.0217	92
Other	21	12,799	.1235	.0188	.0016	.0172	92
Total	127	72,905	.1697	.0178	.0016	.0161	91
			Army				
General Cog	6	888	.1132	.0076	.0066	.0010	13
Personality	4	772	.0799	.0132	.0051	.0071	58
Info Process	0	--	--	--	--	--	--
Job Sample	2	454	--	--	--	--	--
Bio Inventory	0	--	--	--	--	--	--
Psych Coord	8	3,862	.2711	.0023	.0018	.0006	24
Comp/Battery	12	10,476	.1939	.0041	.0011	.0031	74
Other	4	3,492	-.0884	.0143	.0011	.0132	92
Total	36	19,944	.1546	.0208	.0017	.0190	92

Table 11 (Continued)

Predictor Measure	Number of r	Total Sample	Mean r	δ_r^2	δ_e^2	δ_p^2	Percent Unexp.
			Civilian				
General Cog	11	1,567	.2497	.0199	.0062	.0136	69
Personality	5	608	-.0263	.0272	.0083	.0190	70
Info Process	5	577	.3286	.0428	.0070	.0358	84
Job Sample	0	--	--	--	--	--	--
Bio Inventory	1	37	--	--	--	--	--
Psych Coord	3	162	.0214	.0003	.0185	-.0182	0
Comp/Battery	0	--	--	--	--	--	--
Other	2	674	--	--	--	--	--
Total	27	3,625	.1701	.0368	.0071	.0298	81

All combinations with fewer than three correlation coefficients were ignored.

As shown in Table 12, although there is a shift in the order, job sample, psychomotor coordination, and biographical inventory measures are also the three best predictors for the United States subsample, when the data are disaggregated into national groupings. In addition, the variance of the job sample measures decreases substantially, with only 16% of the variance remaining unexplained.

However, while the job sample measures moved to second place for prediction of the United States subsample, they were clearly the best predictors for both the United Kingdom and Canadian subsamples, with mean correlations of .4638 and .3936, respectively. The variances of the job sample correlations differed substantially among the three nations; however, while the unexplained variance for the Canadian subsample was zero, the unexplained variance for the United Kingdom subsample was 91%. This was possibly due to the presence of one United Kingdom study which reported an unusually high validity coefficient for a job sample measure (light-plane screening).

Table 12

Average Validity Coefficients for Predictor-Sample Nationality Combinations

Predictor Measure	Number of r	Total Sample	Mean r	δ_r^2	δ_e^2	δ_p^2	Percent Unexp.
			<u>United States</u>				
General Cog	162	283,243	.1939	.0118	.0006	.0112	95
Personality	39	19,590	.1131	.0152	.0019	.0133	87
Info Process	20	10,146	.2326	.0161	.0018	.0143	89
Job Sample	9	1,734	.2763	.0053	.0044	.0008	16
Bio Inventory	21	27,596	.2663	.0108	.0007	.0101	94
Psych Coord	48	37,286	.3231	.0104	.0010	.0093	90
Comp/Battery	34	35,589	.1934	.0228	.0009	.0219	96
Other	33	33,269	.0912	.0455	.0010	.0445	98
Total	366	403,453	.1977	.0187	.0008	.0179	96
			<u>United Kingdom</u>				
General Cog	6	1,163	.1535	.0091	.0049	.0042	46
Personality	6	852	.1336	.0022	.0068	-.0046	0
Info Process	1	183	--	--	--	--	--
Job Sample	3	226	.4638	.0912	.0082	.0830	91
Bio Inventory	0	--	--	--	--	--	--
Psych Coord	8	1,021	.2240	.0059	.0071	-.0012	0
Comp/Battery	0	--	--	--	--	--	--
Other	0	--	--	--	--	--	--
Total	24	3,445	.1880	.0181	.0065	.0116	64
			<u>Canada</u>				
General Cog	30	5,292	.1617	.0224	.0054	.0170	76
Personality	3	831	-.0286	.0096	.0036	.0060	62
Info Process	6	1,435	.2586	.0379	.0037	.0343	90
Job Sample	4	862	.3936	.0002	.0033	-.0031	0
Bio Inventory	1	366	--	--	--	--	--
Psych Coord	8	957	.0805	.0286	.0083	.0203	71
Comp/Battery	0	--	--	--	--	--	--
Other	0	--	--	--	--	--	--
Total	52	9,743	.1715	.0312	.0051	.0261	84

Table 12 (Continued)

Predictor Measure	Number of r	Total Sample	Mean r	δ_r^2	δ_e^2	δ_p^2	Percent Unexp.
			Other				
General Cog	20	5,514	.1662	.0029	.0034	-.0005	0
Personality	2	2,616	--	--	--	--	--
Info Process	1	1,308	--	--	--	--	--
Job Sample	0	--	--	--	--	--	--
Bio Inventory	0	--	--	--	--	--	--
Psych Coord	9	3,629	.1829	.0036	.0023	.0013	36
Comp/Battery	0	--	--	--	--	--	--
Other	2	2,616	--	--	--	--	--
Total	34	15,683	.1541	.0132	.0021	.0112	84

All combinations with fewer than three correlation coefficients were ignored.

Table 13 compares the validities for fixed-wing (typically Air Force and Navy) and rotary-wing (typically Army) aircraft. For the fixed-wing aircraft the best predictors are the job sample, psychomotor coordination, and biographical inventory measures. Psychomotor coordination is also the best predictor for the rotary-wing subsample, with too little data available to compute validities for the other two measures. The second best predictor for the rotary-wing subsample, in the absence of data for the job sample and biographical inventory measures, is the general cognitive measure, with a mean validity of .1511. Significant here is the very small amount of unexplained variance (2%) for the general cognitive measure category, indicating that sampling error was virtually the sole source of variability among the correlations in that category.

Table 13

Average Validity Coefficients for Predictor-Aircraft Type Combinations

Predictor Measure	Number of r	Total Sample	Mean r	δ_r^2	δ_e^2	δ_p^2	Percent Unexp.
			Fixed-Wing				
General Cog	194	246,426	.1930	.0120	.0007	.0112	94
Personality	46	23,117	.1180	.0204	.0019	.0185	91
Info Process	28	13,072	.2256	.0176	.0019	.0156	89
Job Sample	14	2,368	.3256	.0179	.0047	.0131	74
Bio Inventory	22	27,962	.2646	.0109	.0007	.0102	94
Psych Coord	59	38,065	.3112	.0132	.0013	.0119	90
Comp/Battery	22	25,113	.1932	.0306	.0008	.0298	97
Other	31	32,393	.1080	.0416	.0009	.0407	98
Total	416	408,516	.1998	.0188	.0009	.0179	95
			Rotary-Wing				
General Cog	24	3,786	.1511	.0062	.0061	.0001	2
Personality	4	772	.0799	.0123	.0051	.0071	58
Info Process	0	--	--	--	--	--	--
Job Sample	2	454	--	--	--	--	--
Bio Inventory	0	--	--	--	--	--	--
Psych Coord	14	4,828	.2425	.0066	.0026	.0040	61
Comp/Battery	12	10,476	.1939	.0041	.0011	.0031	74
Other	4	3,492	-.0884	.0143	.0011	.0132	92
Total	60	23,808	.1545	.0184	.0024	.0159	87

All combinations with fewer than three correlation coefficients were ignored.

The final set of comparisons at this level of disaggregation is among the predictor measures for dichotomous and continuous criteria. As before, the same three measures are the best predictors for the dichotomous criterion subgroup. For the continuous criterion subgroup the best predictor is the information processing measure category, followed by psychomotor coordination. Data are not available for the job sample and biographical inventory measures for this criterion group.

Table 14

Average Validity Coefficients for Predictor-Criterion Type Combinations.

Predictor Measure	Number of r	Total Sample	Mean r	δ_r^2	δ_e^2	δ_p^2	Percent Unexp.
			Dichotomous				
General Cog	188	244,031	.1929	.0119	.0007	.0112	94
Personality	43	20,569	.1139	.0141	.0020	.0120	85
Info Process	21	11,087	.2288	.0172	.0017	.0155	90
Job Sample	14	2,692	.3243	.0156	.0042	.0114	73
Bio Inventory	21	27,925	.2646	.0109	.0007	.0102	94
Psych Coord	68	40,115	.3114	.0127	.0014	.0113	89
Comp/Battery	21	25,005	.1939	.0306	.0008	.0298	97
Other	28	28,777	.1147	.0463	.0009	.0454	98
Total	4[illegible]4	400,201	.2021	.0184	.0009	.0175	95
			Continuous				
General Cog	30	6,181	.1707	.0120	.0046	.0074	62
Personality	7	3,320	.1348	.0579	.0020	.0559	96
Info Process	7	1,985	.2075	.0190	.0032	.0158	83
Job Sample	2	130	--	--	--	--	--
Bio Inventory	1	37	--	--	--	--	--
Psych Coord	5	2,778	.1896	.0021	.0017	.0005	22
Comp/Battery	13	10,584	.1922	.0044	.0011	.0032	74
Other	7	7,108	-.0157	.0128	.0010	.0118	92
Total	72	32,123	.1378	.0211	.0022	.0190	90

All combinations with fewer than three correlation coefficients were ignored.

Validities of Specific Predictor Measures

To evaluate the relative validities of specific predictor measures, the general cognitive subgroup was disaggregated into a number of more specific predictor measures. Table 15 contains the mean validity coefficients and variances of those measures, along with two measures (age and education) extracted from the Other subgroup, and three measures (job sample, biographical inventory, and psychomotor coordination) from the General Predictor Category list.

Among these measures, the job sample (r = .3272) remained the best predictor, followed by psychomotor coordination (r = .3035). Next, however, is reaction time, followed by mechanical ability, biographical inventory, general information, aviation information, and perceptual speed--after which the mean correlations slip below .2000. Although there is still a large proportion of unexplained variance for these measures (averaging around 90%), the absolute amount of variance is small in relation to the size of the validities (at least for the larger validities).

The confidence interval for the job sample measure is +/- .2008, making the 95% range for the correlation .1264 to .5280. While this range is still wider than one might like in evaluating the true population correlation, it is safely higher than zero, thus providing assurance that the measures are valid. Toward the bottom of the list, the confidence interval for the aviation information measure is +/- .1828; making the 95% range for the correlation .0496 to .4152.

Table 15

Validity Coefficients for Specific Sets of Measures

Predictor Measure	Number of r	Total Sample	Mean r	δ_r^2	δ_e^2	δ_p^2	Percent Unexp.
General Intellect	12	7,927	.1294	.0078	.0015	.0064	81
Verbal Ability	14	20,756	.1244	.0124	.0007	.0118	95
Quantitative Ability	31	44,799	.1036	.0025	.0007	.0018	72
Spatial Ability	35	47,247	.1851	.0055	.0007	.0048	87
Perceptual Speed	41	29,732	.2001	.0078	.0013	.0066	84
Manual Dexterity	11	2,547	.1044	.0099	.0042	.0057	57
Reaction Time	7	6,854	.2953	.0081	.0009	.0072	89
Mechanical Ability	37	38,708	.2890	.0096	.0008	.0088	92
Aviation Information	18	21,196	.2324	.0094	.0008	.0087	92
General Information	14	27,480	.2536	.0131	.0004	.0126	97
Education *	8	5,495	.0456	.0117	.0015	.0103	88
Age *	8	13,142	-.0964	.0062	.0006	.0056	90
Job Sample **	16	2,822	.3272	.0150	.0045	.0105	70
Bio Inventory **	22	27,962	.2646	.0109	.0007	.0102	94
Psychomotor Coord **	73	42,893	.3035	.0129	.0014	.0115	89

* From Other category
** From General Predictor measures

Discussion and Conclusions

The results of these analyses have shown that three classes of measures are consistent, valid predictors of pilot training performance. Those measures are: job sample, psychomotor coordination, and biographical information. The measures are approximately equally predictive of performance across services and nationalities and (to the extent data are available) in both fixed-wing and rotary-wing aircraft. These findings, therefore support the notions of validity generalization advanced by Schmidt & Hunter (1977, 1981; Schmidt, 1988).

The data have also produced much more stable estimates of the population validities for a number of measures which have been evaluated and/or used for pilot selection over the last 50 years. In the aggregate, the variance of these estimates is still distressingly high, suggesting the need for further research to investigate moderator variables. However, even at this level, the data clearly indicate that the true population correlations are almost certainly not zero.

In addition, because this meta analysis corrected only for sampling error, the estimates of the true validities are conservative. There are other corrections which, while not attempted in this study, could be applied in future research to improve the estimates. Principal among these corrections are (a) correction for unreliability of the criterion, (b) correction for attenuation due to range restriction (which occurs when individuals are selected for entry into training based upon scores on the measure being evaluated), and (c) correction for attenuation due to dichotomization of the criterion (i.e., use of a pass/fail criterion measure). As Hunter & Schmidt (1990) point out, correction factors may be calculated for each of these attenuation effects and applied to the validity coefficients. These have the effect of increasing the validity coefficients by some factor, while at the same time increasing the variance of the estimate.

For the most part, however, the data required to calculate these correction coefficients are missing from the literature. The only relevant datum which is reported with some regularity (but often unintentionally) is the proportions of cases in the pass and fail criterion groups, from which the P/Q split proportions, and hence the correction for dichotomization, may be computed. Those data are available for approximately 90% of the validities in the current study and will be applied in follow-on research. The data for other corrections, such as reliability of the measures or criterion and variances of the unrestricted groups, are uniformly missing. In only a very few cases do the studies report both the uncorrected and corrected correlations for operational selection measures.

Although application of correction factors would increase the estimated correlations, the relative orderings of the validities should remain approximately constant. Even in the lack of these corrections, therefore, we may remain fairly confident regarding which are the best predictors of pilot performance, and which are the worst.

We may conclude, therefore, that the most effective system for the selection of aircraft pilots would include measures of job sample performance, psychomotor coordination and a biographical inventory, along with measures of mechanical ability and reaction time (choice) and such other measures as time and budget allow. We would further conclude that educational attainment has very little relationship to performance in flight training (although many of the military services continue to stress the requirement for a college degree, perhaps to further the professionalism of the officer corps). The contribution of personality measures is also questionable at present, although additional studies which evaluate alternative treatments of the signs of the validities are warranted and might produce better insights into the underlying validities of those measures.

Many other analyses evaluating different aspects of the validities constituting this database are possible and, as questions of interest arise, may be addressed in future research. Certainly, one aspect which will be investigated is the correction for dichotomization of the criterion, for which data in the majority of studies are available. With the growth of interest in the application of meta analytic techniques, one may hope that future studies will report the full set of data required to calculate all applicable correction factors, thus facilitating the development of a high quality pool of research information for future investigation.

References

Hunter, D. R. (1989). Aviator Selection. In M. F. Wiskoff & G. M. Rampton (Eds.), Military Personnel Measurement. New York: Praeger.

Hunter, D. R., & Burke, E. F. (1990). An annotated bibliography of the aircrew selection literature (ARI Research Report 1575). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A230 484)

Hunter, J. E., & Schmidt, F. L. (1990). Methods of meta analysis: Correcting error and bias in research findings. Newbury Park, CA: Sage Publications.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). Meta analysis: Cumulating research findings across studies. Beverly Hills: Sage Publications.

Pearlman, K. (1979). The validity of tests used to select clerical personnel: A comprehensive summary and evaluation (Tech. Study TS-79). Washington, DC: U.S. Office of Personnel Management, Personnel Research and development Center.

Schmidt, F. L. (1988). Validity generalization and the future of criterion-related validity. In H. Wainer & H. Braun (Eds.), Test Validity (pp. 173-189). Hillsdale, NJ: Lawrence Erlbaum.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 65, 643-661.

Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. American Psychologist, 36, 1128-1137.

Schmitt, N., Gooding, R. Z., Noe, R. A., & Kirsch, M. (1984). Meta analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

Siem, F. M., Carretta, T. R., & Mercatante, T. A. (1987). Personality, attitudes, and pilot training performance: Preliminary analysis (AFHRL-TP-87-62). Brooks AFB, TX: Air Force Human Resources Laboratory.

Appendix A

Studies Used in the Meta Analysis

Ambler, R. K., Bair, J. T., & Wherry, R. J. (1960). Factorial structure and validity of naval aviator selector variables. Aerospace Medicine, 31, 456-461.

Ambler, R. K., Johnson, C. W., & Clark, B. (1952). An analysis of biographical inventory and spatial apperception test scores in relation to other selection tests (Special Report 52-5). Pensacola, FL: US Naval School of Aviation Medicine.

Arth, T. O., Steuck, K. W., Sorrentino, C. T., & Burke, E. F. (1988) Air Force Officer Qualifying Test (AFOQT): Predictors of undergraduate pilot training and undergraduate navigator training success (AFHRL-TP-88-27). Brooks AFB, TX: Air Force Human Resources Laboratory.

Bair, J. T., Lockman, R. F., & Martoccia, C. T. (1956).Validity and factor analysis of naval air training predictor and criterion measures. Journal of Applied Psychology, 40, 213-219.

Bale, R. M., & Ambler, R. K. (1971). Application of college and flight background questionnaires as supplementary noncognitive measures for use in the selection of student naval aviators. Aerospace Medicine, 42, 1178-1181.

Bartram, D. & Dale, H. C. A. (1982). The Eysenck Personality Inventory as a selection test for military pilots. Journal of Occupational Psychology, 55, 287-296.

Berkshire, J. R. (1967). Evaluation of several experimental aviation selection tests. Pensacola, FL: Naval Aerospace Medical Center.

Berkshire, J. R., & Ambler, R. K. (1963). The value of indoctrination flights in the screening and training of Naval aviators. Aerospace Medicine, 34, 420-423.

Bordelon, V. P., & Kantor, J. E. (1986) Utilization of psychomotor screening for USAF pilot candidates: Independent and integrated selection methodologies (AFHRL-TR-86-4). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Burke, E. F. (1980). Results of a preliminary study on a new tracking test for pilot selection (Note No. 9/80). London, England: Science 3 (Royal Air Force), Ministry of Defence.

Carretta, T. R. (1987b). The Basic Attributes Tests: An experimental selection and classification instrument for U.S. Air Force pilot candidates. In R.S. Jensen (Ed.), Proceedings of the Fourth International Symposium on Aviation Psychology. Ohio State University: Aviation Psychology Laboratory.

Carretta, T. R., & Siem, F. M. (1988). Personality, attitudes, and pilot training performance: Final analysis (AFHRL-TP-88-23). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Cox, R. H. (1988). Utilization of psychomotor screening for USAF pilot candidates: Enhancing predictive validity. Aviation, Space, and Environmental Medicine, 59, 640-645.

Croll, P. R., Mullins, C. J., & Weeks, J. L. (1973). Validation of the cross-cultural aircrew aptitude battery on a Vietnamese pilot trainee sample (AFHRL-TR-73-30). Brooks Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Damos, D. L., & Lintern, G. (1979). A comparison of single- and dual-task measures to predict pilot performance (Technical Report Eng Psy-79/2). Urbana-Champaign, IL: University of Illinois at Urbana-Champaign.

Davis, R. A. (1989). Personality: Its use in selecting candidates for US Air Force undergraduate pilot training (Research Report No. AU-ARI-88-8). Maxwell Air Force Base, AL: Air University Press.

DeWet, D. R. (1963). The roundabout: a rotary pursuit-test, and its investigation on prospective air-pilots. Psychologia Africana, 10, 48-62.

Doll, R. E. (1962). Officer peer ratings as a predictor of failure to complete flight training (Special Report 62-2). Pensacola, FL: U. S. Naval Aviation Medical Center.

Elshaw, C. C., & Lidderdale, I. G. (1982). Flying selection in the Royal Air Force. Revue de Psychologie Appliquie (Supplement), 32, 3-13.
[Alternative citation is: Newsletter of the International Test Commission of the Division of Psychological Assessment of the International Association of Aviation Psychologists, 17, December 1982]

Fiske, D. W. (1947). Validation of naval aviation cadet selection tests against training criteria. Journal of Applied Psychology, 31, 601-614.

Flanagan, J. C. (1947). The aviation psychology program in the Army Air Forces. AAF Aviation Psychology Program Research Report No. 1. Washington, DC: U. S. Government Printing Office.

Fleischman, H. L., Ambler, R. K., Peterson, F. E., & Lane, N. E. (1966). The relationship of five personality scales to success in naval aviation training (NAMI - 968). Pensacola, FL: Naval Aerospace Medical Institute.

Fleishman, E. A. (1954). Evaluations of psychomotor tests for pilot selection: the direction control and compensatory balance tests (AFPTRC-TR-54-131). Lackland Air Force Base, TX: Air Force Personnel & Training Research Center.

Fleishman, E. A. (1956). Psychomotor selection tests: research and application in the United States Air Force. Personnel Psychology, 9, 449-467.

Flyer, E. S., & Bigbee, L. R. (1954). The light plane as a pre-primary selection and training device: III. Analysis of selection data (AFPTRC-TR-54-125). Lackland Air Force Base, TX: Air Force Personnel & Training Research Center.

Fowler, B. (1981). The aircraft landing test: an information processing approach to pilot selection. Human Factors, 23, 129-137.

Goebel, R. A., Baum, D. R., & Hagin, W. V. (1971). Using a ground trainer in a job sample approach to predicting pilot performance (AFHRL-TR-71-50). Williams Air Force Base, AZ: Flying Training Division, Air Force Human Resources Laboratory.

Gopher, D. (1982). A selective attention test as a predictor of success in flight training. Human Factors, 24, 173-183.

Gopher, D., & Kahneman, D. (1971). Individual differences in attention and the prediction of flight criteria. Perceptual and Motor Skills, 33, 1335-1342.

Gordon T. (1949). The airline pilot's jobs. Journal of Applied Psychology, 33, 122-131.

Graybiel, A., & West, H. (1945). The relationship between physical fitness and success in training of U. S. Naval flight students. Journal of Aviation Medicine, 16, 242-249.

Greene, R. R. (1947) Studies in pilot selection. II. The ability to perceive and react differentially to configuration changes as related to the piloting of light aircraft. Psychological Monographs, 61, 18-28.

Griffin, G. R., & McBride, D. K. (1986). Multitask performance: predicting success in naval aviation primary flight training (NAMRL-1316). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Griffin, G. R., & Mosko, J. D. (1982). Preliminary evaluation of two dichotic listening tasks as predictors of performance in naval aviation undergraduate pilot training (NAMRL-1287). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Guilford, J. P., & Lacey, J. I. (1947). Printed classification tests. AAF Aviation Psychology Program Research Report No. 5. Washington, DC: U. S. Government Printing Office.

Guinn, N., Vitola, B. M., & Leisey, S. A. (1976). Background and interest measures as predictors of success in undergraduate pilot training (AFHRL-TR-76-9). Lackland Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Hertli, P. (1982). The prediction of success in Army aviator training: A study of the warrant officer candidate selection process (Unpublished Report). Fort Rucker, AL: U. S. Army Research Institute Field Unit.

Hunter, D. R. (1982). Air Force pilot selection research. Paper presented at the 90th Annual Meeting of the American Psychological Association. Washington, DC.

Hunter, D. R., & Thompson, N. A. (1978). Pilot selection system development (AFHRL-TR-78-33). Brooks Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Joaquin, J. B. (1980b). The Personality Research Form (PRF) and its utility in predicting undergraduate pilot training performance in the Canadian Forces (Working Paper 80-12). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

Kaplan, H. (1965). Prediction of success in Army aviation training (Technical Research Report 1142). Washington, DC: U.S. Army Personnel Research Office. (AD A623 046)

King, J. E. (1945). Relation of aptitude tests to success of Negro trainees in elementary pilot training (Research Bulletin 45-52). Tuskegee Army Air Field: Office of the Surgeon, Headquarters Army Air Forces Training Command.

Knight, S. (1978). Validation of RAF pilot selection measures (Note No. 7/78). London, England: Science 3 (Royal Air Force), Ministry of Defence.

Koonce, J. M. (1981). Validation of a proposed pilot trainee selection system. In R. S. Jensen (Ed.), Proceedings of the First Symposium on Aviation Psychology (Technical Report APL-1-81). Columbus, OH: Aviation Psychology Laboratory of the Ohio State University.

Lane, G. G. (1947). Studies in pilot selection: I. The prediction of success in learning to fly light aircraft. Psychological Monographs, 61, 1-17.

LeMaster, W. D., & Gray, T. H. (1974). Ground training devices in job sample approach to UPT selection and screening (AFHRL-TR-74-86). Williams Air Force Base, AZ: Flying Training Division, Air Force Human Resources Laboratory.

Lidderdale, I. G. (1976). The primary flying grading trial interim report No. 2. RAF Brampton, England: Research Branch, Headquarters Command, Royal Air Force, Ministry of Defence.

McAnulty, D. M. (1990). Validation of an experimental battery of Army aviator ability tests. Fort Rucker, AL: Anacapa Sciences; Inc.

McGrevy, D. F., & Valentine, L. D. (1974). Validation of two aircrew psychomotor tests (AFHRL-TR-74-4). Lackland Air Force Base, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Melton, A. W. (Ed.) (1947). Army Air Forces Aviation Psychology Research Reports: Apparatus Tests (Report No. 4). Washington, DC: U. S. Government Printing Office.

Miller, J. T., Eschenbrenner, A. J., Marco, R. A., & Dohme, J. A. (1981). Mission track selection process for the Army initial entry rotary wing flight training program. St. Louis, MO: McDonnel Douglas Astronautics Co.

Mullins, C. J., Keeth, J. B., & Riederich, L. D. (1968). Selection of foreign students for training in the United States Air Force (AFHRL-TR-68-111). Lackland Air Force Base, TX: Personnel Research Division, Air Force Human Resource Laboratory.

North, R. A., & Gopher, D. (1974). Basic attention measures as predictors of success in flight training (Technical Report ARL-74-14). Urbana-Champaign, IL: Aviation Research Laboratory, University of Illinois at Urbana-Champaign.

Owens, J. M., & Goodman, L. S. (1983). Navy aviation selection and classification research. Paper presented at the Eleventh Meeting of the Department of Defense Human Factors Engineering Technical Advisory Group, Aberdeen Proving Grounds, MD.

Roth, J. T. (1980). Continuation of data collection on causes of attrition in initial entry rotary wing training. Valencia, PA: Applied Science Associates.

Sells, S. B. (1956). Further developments on adaptability screening of flying personnel. Journal of Aviation Medicine, [illegible] 440-451.

Sells, S. B., [illegible]rites, D. K., Templeton, R. C., & Seaquist, M. R. (1958). Adaptability screening of flying personnel: cross validation of the personal history blank under field conditions. Washington, DC: Proceedings of the 29th Annual Meeting of the Aero Medical Association.

Shipley, B. D. (1983). Maintenance of level flight in a UH-1 flight simulator as a predictor of success in Army flight training. Unpublished manuscript. Fort Rucker, AL: United States Army Research Institute.

Shoenberger, R. W., Wherry, R. J., & Berkshire, J. R.(1963). Predicting success in aviation training (Report No. 7). Pensacola, FL: U. S. Naval School of Aviation Medicine.

Shull, R. N., & Dolgin, D. L. (1989). Personality and flight training performance. In Proceedings of the Human Factors Society--33rd Annual Meeting.

Shull, R. N., Dolgin, D. L., & Gibb, G. D. (1988). The relationship between flight training performance, a risk assessment test, and the Jenkins Activity Survey. (NAMRL-1339). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Siem, F. M. (1988). Personality characteristics of USAF pilot candidates. Proceedings of the 1988 AGARD Meeting on Aircrew Performance. Paris:

Signori, E. I. (1949). The Arnprior Experiment: a study of World War II pilot selection procedures in the RCAF and RAF. Canadian Journal of Psychology, 3, 136-150.

Stoker, P. (1982). An empirical investigation of the predictive validity of the defence mechanism test in the screening of fast-jet pilots for the Royal Air Force. Projective Psychology, 27, 7-12.

Stoker, P., Hunter, D. R., Kantor, J. E., Quebe, J. C., & Siem, F. M. (1987). Flight screening program effects on attrition in undergraduate pilot training (AFHRL-TP-86-59). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Trankell, A. (1959). The psychologist as an instrument of prediction. Journal of Applied Psychology, 43, 170-175.

Tucker, J. A. (1954). Use of previous flying experience as a predictor variable (AFPTRC-TR-54-71). Lackland Air Force Base, TX: Air Force Personnel & Training Research Center.

Voas, R. B. (1959). Vocational interests of naval aviation cadets: final results. Journal of Applied Psychology, 43, 70-73.

Want, R. L. (1962). The validity of tests in the selection of Air Force pilots. Australian Journal of Psychology, 14, 133-139.

Appendix B

Meta Analysis Computer Program

[These commands are stored in a separate file called "RUNMETA.PRG", and define the records to be selected for analysis. It invokes a separate procedure file called "META.PRG" to perform the calculations.]

```
USE metadat4          [metadat4 is the name of the database file]
GO TOP
STORE 0.0 TO VAR08
SET DEVICE TO PRINT
SET ECHO OFF
SET FILTER TO
STORE 'ALL' TO VAR09  [This example run first uses All studies]
COUNT TO VAR08
DO META
SET FILTER TO TEST_CAT = "A"   [Next, it uses each of the
                        measure categories separately]
STORE 'TEST CATEGORY = A (General Cognitive)' TO VAR09
COUNT FOR TEST_CAT = "A" TO VAR08
DO META
SET FILTER TO TEST_CAT = "B"
STORE 'TEST CATEGORY = B (Personality)' TO VAR09
COUNT FOR TEST_CAT = "B" TO VAR08
DO META
SET FILTER TO TEST_CAT = "C"
STORE 'TEST CATEGORY = C (Info Processing)' TO VAR09
COUNT FOR TEST_CAT = "C" TO VAR08
DO META
SET FILTER TO TEST_CAT = "D"
STORE 'TEST CATEGORY = D (Job Sample)' TO VAR09
COUNT FOR TEST_CAT = "D" TO VAR08
DO META
SET FILTER TO TEST_CAT = "E"
STORE 'TEST CATEGORY = E (Other)' TO VAR09
COUNT FOR TEST_CAT = "E" TO VAR08
DO META
SET FILTER TO TEST_CAT = "F"
STORE 'TEST CATEGORY = F (Biographical Inventories)' TO VAR09
COUNT FOR TEST_CAT = "F" TO VAR08
DO META
SET FILTER TO TEST_CAT = "G"
STORE 'TEST CATEGORY = G (Psychomotor Coordination)' TO VAR09
COUNT FOR TEST_CAT = "G" TO VAR08
DO META
SET FILTER TO TEST_CAT = "X"
STORE 'TEST CATEGORY = X (Batteries or composites)' TO VAR09
COUNT FOR TEST_CAT = "X" TO VAR08
DO META
```

[This is the meta analysis procedure file. The following code is stored in a separate file called "META.PRG" and performs the meta analysis calculations using the records selected by "RUNMETA.PRG".]

```
GO TOP
SET DECIMALS TO 4
STORE 0.00 TO VAR01, VAR02, VAR03, VAR04, VAR05, VAR06, VAR07
STORE 0.00 TO VAR10, VAR11, VAR12, VAR13
CLEAR
SUM (SAMPLE_N * CORRELATN) TO VAR01  && Sum of weighted r's
SUM SAMPLE_N TO VAR02                && Total sample size
STORE VAR01 / VAR02 TO VAR03         && Mean weighted r
SUM (SAMPLE_N * (CORRELATN - VAR03)**2) TO VAR04
STORE VAR04 / VAR02 TO VAR05         && Total Variance
STORE (1 - (VAR03)**2)**2 /(VAR02/VAR08 - 1) TO VAR06 && var(e)
STORE VAR05 - VAR06 TO VAR07         && True Variance
STORE (VAR07 / VAR05) * 100 TO VAR10 && % Var unaccounted
STORE SQRT(VAR07) TO VAR11           && Corrected S.D. of r's
STORE VAR03 + 1.96 * VAR11 TO VAR12  && Upper confidence bound
STORE VAR03 - 1.96 * VAR11 TO VAR13  && Lower confidence bound
CLEAR
@  1,0 SAY'****************************************************'
@  3,20 SAY 'Meta Analysis Program'
@  4,22 SAY 'Version 1.2'
@  5,15 SAY 'Correction for Sampling Errors'
@  6,5 SAY '____________________________________________________'
@  8,5  SAY 'Records selected:'
@  8,25 SAY VAR09
@ 10,5 SAY 'The number of correlations cumulated (k) is:'
@ 10,45 SAY VAR08
@ 11,5 SAY 'The total sample (N) is:'
@ 11,45 SAY VAR02
@ 12,5 SAY 'The average weighted r is: '
@ 12,45 SAY VAR03
@ 13,5 SAY 'The Total Variance is:     '
@ 13,45 SAY VAR05
@ 14,5 SAY 'The Error Variance is:    '
@ 14,45 SAY VAR06
@ 15,5 SAY 'The True (corrected) Variance is:   '
@ 15,45 SAY VAR07
@ 16,5 SAY 'The Percentage of Unexplained Variance is: '
@ 16,45 SAY VAR10
@ 17,5 SAY 'The Standard Deviation (corrected) for r is: '
@ 17,45 SAY VAR11
@ 18,5 SAY 'The Upper Confidence Bound (r + 1.96 * SD) is:'
@ 18,45 SAY VAR12
@ 19,5 SAY 'The Lower Confidence Bound (r - 1.96 * SD) is:'
@ 19,45 SAY VAR13
@ 21,0  SAY '***************************************************'
@ 22,0 SAY ''
```